

# COMPUTER CENTRE BULLETIN

Vol 5 No 5  
8 May 1972

Editor:  
Mrs Sarah Barry



WITHDRAWAL OF THE GE-225 BY 1975

The initial GE-225 computer was installed in the Centre in March 1962 and since that date the machine configuration has been expanded to its present level.

As the GE-225 machine is now 10 years old and the Centre has experienced difficulty in obtaining some of the spare parts needed to maintain the equipment in good operational condition, it will be impractical and uneconomic to provide GE-225 services beyond 1975.

The University therefore resolved:

"that the GE-225 computer services remain available during 1973-75, providing finance is available for this machine to be maintained in a reliable condition, and, in any case, use of this machine should terminate at the end of 1975."

Thus, provided that both adequate finance and supplies of spare parts continue to be available, the Centre intends to provide GE-225 services until 24 December 1975. No GE-225 services will be available beyond that date.

To guard against the possibility of an earlier termination of the GE-225 system, users are advised to convert their applications to the PDP-10 as soon as practicable.

RUNOFF

[WN-83]

A new version of Runoff (V7(21)-2) has become operational. This version incorporates some new U.Q. commands which are documented in the Runoff manual, MNT-14. While these do not affect users currently preparing documentation with Runoff, it does affect the RUNOFF command slightly.

B5-5  
2 May 72

After Runoff has been invoked, a new question is asked

UQ-STYLE OUTPUT (Y OR N)?

The U.Q. format is regarded as standard and users who do not require it should answer N to the above question.

The Runoff manual is now available at the Computer Centre. It costs \$1.00.

#### NEW PLOTTER SOFTWARE AND MANUAL

[WN-84]

The new plotter manual (MNT-11) price \$2.80, is now available from the Centre. This manual details new basic plotter subroutines for the PDP-10 (which have some differences from those already in use) as well as an extensive suite of functional subroutines and some plotting programs. All these routines are now available on the PLOT directory. Instructions on the use of these routines and how to access them are given on pages 2-28, 3-75 and 4-31 of the plotting manual. However, until the library directories on the PDP-10 system are reorganized, users will have to access library files on the PLOT directory by '\$PLOT.filename' rather than 'PLOT.filename' as specified in the manual. For instance, in the example on page 3-75, the user must say

```
.RUN MAINP/REL $PLOT.CALCMP(LIB)<cr>
```

This is only a temporary change and will be removed in due course.

As the new basic plotter routines have expanded capabilities, users wishing to use the new routines in any of their existing plotting programs may have to modify some of the subroutine calling sequences. The old plotter routines have been part of LIB40, the FORTRAN library, and so were automatically loaded with a user's plotting program by the command

```
RUN program-name
```

As part of the library reorganization these old routines are now available in the file DEC/REL on the PLOT directory. Users who do not wish to change their programs can continue using the old routines simply by changing their run command to

RUN program-name \$PLOT.DEC(LIB)

To provide a change over period, the old plotter routines will remain in LIB40 until 1 July.

At present the old basic subroutines receive type 1 support. With the release of the new plotter software, the new basic subroutines will be given type 1 support; the old basic subroutines and the new functional routines and programs will attract type 3 support.

### GRAPHIC ART COMPETITION

The following article has been received from the Computer Club. Readers of the Bulletin may be interested in entering this competition.

#### 1. COMPETITION

The Computer Club, maintaining its aims with regard to computing on campus, is promoting a Graphic Art Competition. Many people are using the Computer Centre's plotter to artistic advantage and we know there are many others as well, who would want to enter their works in such a competition. For this reason we are placing as few restrictions as possible on entry, and people outside the University are especially welcome to enter.

#### 2. SEMINAR

To assist prospective entrants in preparing their work, the Club is holding a seminar on the UQ plotter software. This will be a lunch hour seminar on Thursday 11 May at 1.10 p.m. in B18 Engineering building.

At this seminar you will learn the operation and use of the plotting facilities on the PDP-10 and this will help you in preparing your entry.

3. COSTS

As a Club, we realize our members may spend a fair sum on a high quality entry. To help offset such cost the Club plans to hold an exhibition of the works, where they will all be offered for sale, at a price nominated by the artist. The Club will retain a small commission of 10%. In addition, prizes to be announced, will be offered.

4. EXAMPLES

To get some idea of what can be done, look at the display now in the clients' room at the Centre. (A sample drawing is shown below).

5. JUDGES

The judging panel of three, will include people prominent in the artistic world and also in computing.

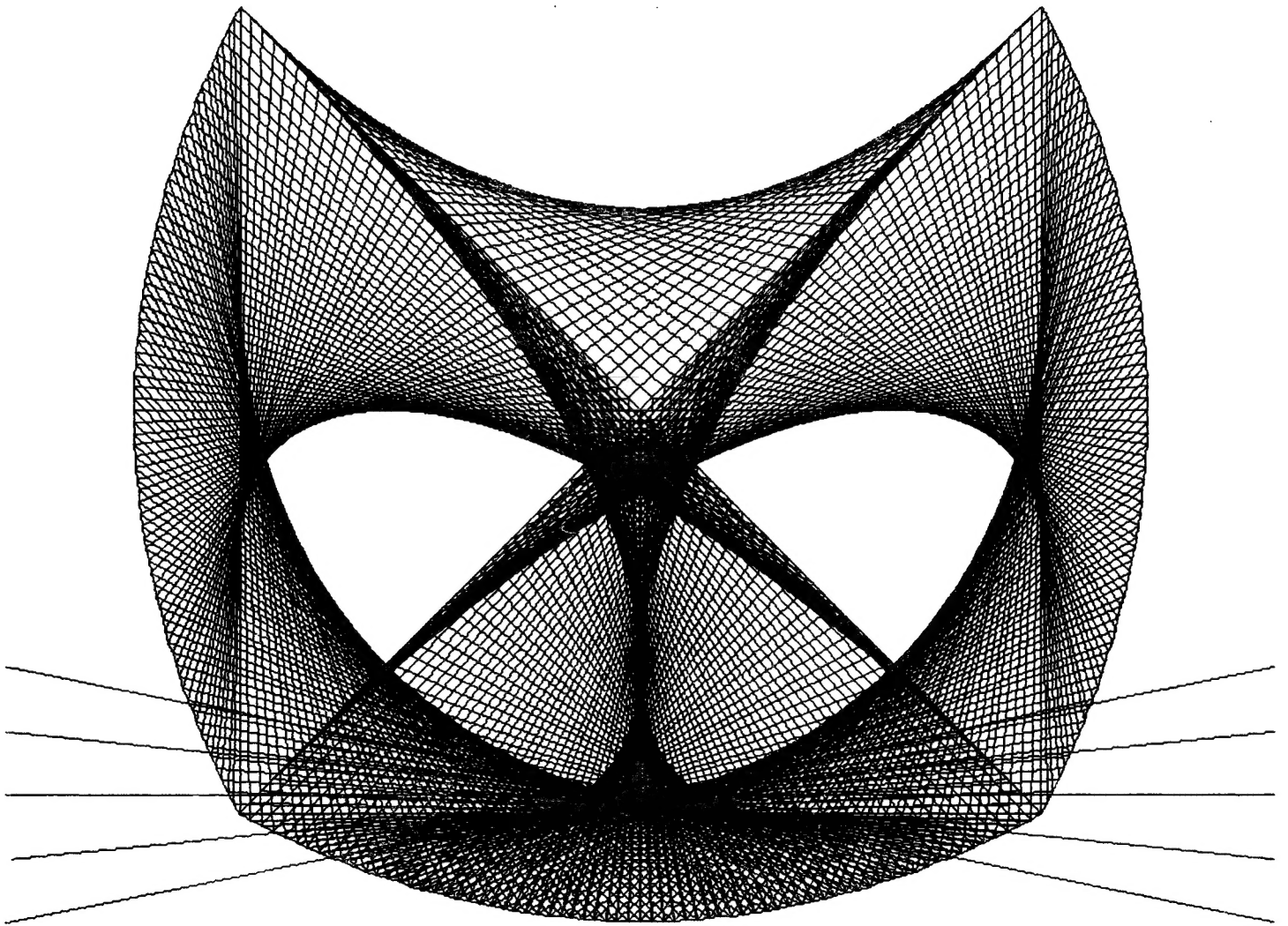
6. RULES

Please read the following rules:

- (a) All members of the Computer Club, and any other interested people, except as hereafter excluded, are able to enter the Club's graphic art competition.
- (b) Personnel of the Computer Centre may not enter the competition.
- (c) Entry fees shall be, for each work entered
  - club members : .50c
  - non members : \$1.00
- (d) All work submitted must be original, suitably titled, and signed by the artist. Each entry must be in black ink on plain white paper.
- (e) After the competition, all works will be offered for sale, at a price nominated by the artist. The Club will keep 10% commission on sales made.
- (f) The maximum size for each entry (i.e. the area bounded by the frame of the work) will be eleven inches by fifteen

inches.

- (g) The decision of the judging panel shall be final and no correspondence regarding the competition's results will be entered into.
- (h) All entries must be in the hands of the club by 5 p.m. on 1 September 1972.



PUSSY

B5-5  
2 May 72

### MAIL

[WN-84]

The MAIL facility on the PDP-10 allows for the interchange of short messages between users. The mail facility is now operational and a detailed write up is available in the March Bulletin, B5-3.

### GLOB

[WN-84]

A new version of GLOB was implemented on 20 April 1972.

### LIBRARY ACCESSIONS

The following books were added to libraries in the University during November, January and February.

FISHER, Gene Harvey	Cost considerations in systems analysis. 1971 (658.151 FIS Engin.)
	Computers and old English concordances. 1970 (PR171.C6 Main)
MIELE, Angelo	Mathematical programming for constrained minimal problems. 1969 (QtoQA264.M5 Maths)
SCHEID, Francis	Schaum's outline of theory and problems of introduction to computer science. 1970 (QtoQA76.S34 Maths)
HOWE, Robert Milton	Design fundamentals of analog computer components. 1961 (510.7824 HOW Elect.)
SUPER, Donald Edwin, comp.	Computer-assisted counseling. 1970 (371.40285 SUP Rem. Ed.)



HARRISON, Malcolm C.	Data-structures and programming 1970 (QtoQA264.H35 Maths)
INTERNATIONAL COMPUTERS LIMITED	Programming the 1900 series in COBOL. 1969 (QtoQA76.I19I5 Main)
PRICE, Wilson T.	Elements of basic FORTRAN IV programming. 1969 (Qto651.8 PRI Engin.)
WOLFENDALE, E.	Computer-aided design techniques. 1970 (621.38153 WOL Elect.)
WORTHINGTON, Henry W.	A computer program to estimate the combined effect of refraction and diffraction of water waves. 1970 (Qto627.12593 WOR Engin.)
NORTH AMERICAN PALEONTOLOGICAL CONVENTION, Field Museum of Natural History, Chicago.	Proceedings. Part B: Computers in paleontology. 1970 (560 NOR Geol.)
SANDERSON, Peter C.	Computers for management. 1969 (HF5548.2.S232 U/G)
LOVIS, Frank Brian	Computers 1. 1964 (510.7834 LOV Educ.)
LOVIS, Frank Brian	Computers 2. 1964 (510.7834 LOV Educ.)
NEW SOUTH WALES UNIVERSITY, Division of Post Graduate Extension Studies	Introduction to matrices and mathematical programmes. 1971 (QtoQA263.N4 Main)
POLACK, Elijah	Computational methods in optimization. 1971 (QA402.5.P58 Maths)
WHITBY, Lionel Gordon, ed.	Principles and practice of medical computing. 1971 (610.2854 WHI PAH)
CLIFTON, Harold Dennis	Data processing systems design. 1971 (T57.5.C54 Main)
KARBOWIAK, Antoni Emil, ed.	Information, computers, machines and man. 1971 (TA168.K3 Main)
LOCKYER, Keith Gerald	An introduction to critical path analysis. 1969 (T57.85.L59 Main)

STEPHENSON, Robert E.

# A PROGRAM FOR CURVE FITTING AND DATA ADJUSTMENT

M. J. Lees

## 1. ABSTRACT

A general purpose FORTRAN curve fitting program has been developed. The program allows the least squares fitting of multi-variate data subject to specified constraints containing unknown parameters. Observations on all variables are assumed to have been drawn from populations with known normally distributed measurement error. Up to ten types of constraints (or curves) may be fitted simultaneously. The curves may be non-linear.

## 2. INTRODUCTION

The algorithm for the curve fitting program has been based on the general least squares solution discussed by Deming [1]. The most important feature of the technique is that in the general least squares problem, all variables are assumed to contribute to the sum of the squares of residuals. This is in contrast to the usual regression problem in which only the dependent variable is assumed to be subject to error. In the curve fitting program discussed in this paper, all variables are treated in similar manner.

The curve fitting procedure minimizes the weighted sum of squares of the residuals obtained for each of the variables at each data point. The weights are based on the relative measurement precisions of individual observations in the data.

The curve fitting simplification given by Deming has been extended [2] to give the solution to the general curve fitting problem in which more than one type of constraint (or curve) is to be fitted.

Mr M. J. Lees is a post-graduate student at the Julius Kruttschnitt Mineral Research Centre.

The flexibility of the technique, which assumes all variables have measurement errors and allows several types of constraints to be specified simultaneously, has proved to be an important feature when used for the analysis of data from mineral processing operations.

### 3. IMPLEMENTATION

The difficulty in implementing the general curve fitting program on a digital computer arises from the large core memory storage required to solve the normal equations (the program was originally written for the GE-225). For a problem in which there are  $n$  data observations,  $u$  parameters and  $r$  types of constraint (or curves), then the matrix of the normal equations contains

$$[(r*n) + u+1]^2$$

elements. By applying the curve fitting simplifications and a suitable reorganization of the normal equations [2], the size of the matrix is reduced to

$$(r+u+1)^2$$

elements. This is achieved by a looping procedure which uses the same core storage  $n$  times. Since the coefficient matrix is symmetric and also contains some null elements, storage need only be provided for

$$r*(r+1)/2 + 2*r*(u+1)$$

elements. The reduction in storage requirements that is achieved is considerable since  $n$  is usually much greater than  $r$  or  $u$ .

The reorganized solution has been programmed in FORTRAN IV. The solution of the equations has been carried out by performing a Cholesky reduction on the lower half of the relevant matrix. The main advantages of this method are that less core storage is required (lower half of matrix stored only) and less operations are required in the computation.

Non-linear problems have been accommodated by simple iteration. A check on the calculated sum of squares ensures that this value does not increase. A predicted increase in the sum of squares at any particular iteration causes the program to step back on the parameter values towards the values held at the end of the preceding iteration. If a lower sum of squares is obtained the solution proceeds, otherwise an accuracy limit message results. The satisfactory solution of many non-linear problems depends on the user providing a sufficiently accurate initial estimate for

the parameters. For linear problems almost any initial parameter estimates are satisfactory.

The program normally terminates at a fixed convergence ratio on the parameter values in the solution ( $10^{-4}$  in this case). However, should this condition not be attained, and the calculated sum of squares was still decreasing at the last step, a non-convergence message is printed.

The derivatives required for the solution are obtained numerically in a separate sub-program. If the analytic values of the derivatives can be easily specified for the curve being fitted, then some increase in speed and accuracy may be obtained by replacing this sub-program by a special version giving the analytic values. However, the gain is usually not very significant.

#### 4. PROGRAM OUTPUT

A detailed summary of the program output is given later in this article. One of the more useful results obtained from the technique is a set of statistically adjusted data values which satisfy the constraint equations specified by the user. The adjustment is carried out according to the manner in which the weighted residuals contribute to the sum of squares. This is the more general representation of the adjustment procedure which corresponds to the adjustment of the dependent variable in regression by the amount given in the table of residuals, i.e. in either case the adjusted data values lie on the calculated curve.

Because the algorithm is used under the assumption that the residuals are small (first order Taylor series approximation), the accuracy of the adjustment is dependent to some extent on the size of the residuals. It is also dependent on the accuracy of the derivatives and the numerical accuracy obtained in the calculations.

As a check on the possibility of obtaining poor data adjustment, the constraint residuals are calculated and printed out both before and after data adjustment. Following data adjustment, the constraint residuals would, in the ideal case, be zero, i.e. the calculated points lie exactly on the calculated curve. However, for practical purposes it is usually sufficient that they be relatively small when compared to the values prior to adjustment.

## 5. EXAMPLE OF USAGE

One application [3] of the program in mineral processing is the calculation of flowrates of pulp from data obtained by sampling pulp streams in industrial installations. This method is necessary if flowrate measuring instruments are not installed. In plant installations where instruments are available, the technique has been used to calibrate the instruments. In this type of problem the parameters represent the unknown flowrates and calibration factors, while the constraints are the relationships defining the network of pulp streams in the plant.

## 6. USING THE PROGRAM

The program is available for usage under the filename CURFIT/REL on project 445. The form of the curve (or curves) to be fitted is specified by the user in the subroutine CURV. The user must also provide a subroutine TRANS. The order of the input data for the main program and the descriptions of CURV and TRANS are given below.

### 6.1 Curve Fitting Main Program (PDP-10)

A general purpose main program for use with the curve fitting subroutine CFITM.

#### (a) Variables in unlabelled common

integers NRL, NINT  
integer array INDAT  
real array RL DAT

#### (b) Input

Data is read from cards in the following order

1. heading
2. date
3. number of variables, parameters, constraint equations, sets of data, iterations, real constants, integer constants.  
FORMAT (15I5)

4. the five character constraint labels  
FORMAT ((7X,A5,9(2X,A5)))
5. the five character variable labels (as for card 4)
6. the first data point label and observed values of  
the variables at that point  
FORMAT (A5,10F7.0/(F12.0,9F7.0))
7. the weights label and the weights for the first data  
point  
(format as for card 6)
- 
- 
- 
- (n+5) last label and nth data observation
- (n+6) last weight label and nth set of weights
- (n+7) the parameter labels (as for card 4)
- (n+8) the parameter estimates  
FORMAT (5X,10F7.0)
- (n+9) the values of the real constants (more than one card  
may be used if necessary)  
FORMAT (5X,10F7.0)
- (n+10) the values of the integer constants (more than one  
card may be used if necessary)  
FORMAT (15I5)
- (n+11) the characters FORCE in columns 1-5 will force the  
adjustment of the data
- (n+12) .END in columns 1-4 indicates the end of the last  
set of data.

(c) Output on the line printer

1. the input data
2. the number of derivative evaluations required during  
the fitting, and the total number of constraint  
points.
3. the fitted parameters

4. the standard errors in the parameters
5. the pooled standard error; this is the minimized standard error and is obtained by pooling the standard error from each of the variables.
6. the constraint residuals obtained using the calculated parameters and observed data
7. the sum of weighted squares for each of the variables
8. the adjusted data
9. the constraint residuals obtained using the calculated parameters and the adjusted data.

(d) Restrictions

A maximum of

- (i) 100 data points
- (ii) 20 variables
- (iii) 10 parameters
- (iv) 10 types of constraint

The output is printed to four digits accuracy.

(e) Subroutines used

HEAD, TABHD, CFITM, CFITMD, CURV, CHECK, TRANS, CFITMR

The subroutine CURV specifies the constraints and is supplied by the user. The subroutine TRANS specifies any transformations on the data and is also supplied by the user.

(f) Remarks

The weights used in the input data are the reciprocals of the standard errors of measurement of the variable at a particular data point.

$$\text{i.e. } w_{ij} = 1/\sigma_{x_{ij}}$$

is the weight of the  $i$ th variable at the  $j$ th data point.

These weights are squared in the program CFITM before they enter the solution. The standard error is used for convenience at data input, because it helps avoid very large (or very small) data weights.

## 6.2 Transform Input Data (PDP-10)

SUBROUTINE TRANS (NV, NS, VNAME, X, W)

A user supplied subroutine for carrying out transformations on the user's data.

### (a) Arguments

NV        number of variables  
NS        number of sets of data  
VNAME     the variable names  
X         array containing NS observations on NV variables  
W         the weights for the data in array X

### (b) Input values

integers NV, NS  
real arrays VNAME, X, W

### (c) Output

integer NV  
real arrays VNAME, X, W

### (d) Side effects

Transformations on the array X must be accompanied by the corresponding transformation on the weight array W.

The elements of W are defined by

$$w_{ij} = 1/\sigma_{x_{ij}}$$

In carrying out transformations on W the following variance relationships may be useful

$$\begin{aligned}\text{Var}(aX) &= a^2 \text{Var}(X) \\ \text{Var}(a+X) &= \text{Var}(X)\end{aligned}$$



(e) Restrictions

New variables produced by transformation must allow the problem size to remain within the maximum allowable limits. If the limit is exceeded during transformation this will be detected by the main program in a call to subroutine CHECK, and a call to EXIT will result.

The arrays should have dimensions  
VNAME(20), X(100,20), W(100,20)

6.3 User Supplied Function (PDP-10)

SUBROUTINE CURV (NE, IS, X, ERR)

User supplied function for calculating the constraint residuals at a point.

(a) Arguments

NE        number of constraint types

IS        number of the data observation being considered

X        array of NV+NP elements. The first NV values are the values of the variables for the IS observation. (NV = number of variables).

The next NP values are the current values of the parameters (NP = number of parameters).

ERR       array of the constraint residuals (observed - calculated) at the IS data point.

(b) In unlabelled common

(see 6.1 main program, also section (e) below)

(c) Output

integer NE (if defined in CURV)  
real array ERR

(d) Restrictions

The arrays should have dimensions X(30), ERR(10)

There is a maximum of 10 types of constraint, 20 variables, and 10 parameters.

(e) Method

The residuals for each of the constraint types is calculated and the answers stored in the first NE elements of ERR.

The value of NE may also be defined from this program instead of from the main program.

Integer and real constants read in from the data into the arrays RLDAT and INDAT may be accessed by an unlabelled COMMON statement of the form

COMMON NRL, RLDAT(50), NINT, INDAT(50)

7. REFERENCES

- [1] Deming, W. E., 1964  
Statistical Adjustment of Data. Dover Publications Inc., New York.
- [2] Lees, M. J., 1971  
A Method for Curve Fitting and Data Adjustment. Julius Kruttschnitt Mineral Research Centre Technical Report. 1st July - 31st December, 1971.
- [3] Lees, M. J. and McKee, D. J., 1971  
An Analysis of Plant Flotation Data Using Curve Fitting Methods, Proc. Symp. on Automatic Control Systems in Mineral Processing Plants, Aus. I.M.M., Brisbane, May 1971.